Soybeans are big business in Missouri, which ranks in the nation’s top five soybean-producing states. In 2014, the tiny beans represented a $4.5 billion dollar industry for the Show-Me State.

The University of Missouri expanded its investment in soybean research in 1999 by hiring four faculty members with expertise in breeding, genetics, genomics and economics to fill endowed chair positions funded by the Missouri Soybean Merchandising Council (MSMC) with assets that came directly from the state’s soybean growers.

Two of the MU hires, Endowed Professor of Genetics and Biotechnology Henry Nguyen and Gary Stacey, a Curators Professor in the Division of Plant Sciences, were named director and associate director, respectively of the university’s National Center for Soybean Biotechnology, launched in 2004. And one of the center’s first orders of business was to sequence the soybean genome.

“Plant science cannot be just biology...
anymore. You have to have the informatics,” Nguyen explained. “We had no bioinformatics capabilities, and [a campus taskforce] spent a year meeting about it. In the process, we recruited Dong Xu, who was familiar with the landscape.”

Xu, now the James C. Dowell Professor and chair of engineering’s Computer Science Department, came to MU with a background in protein structure prediction and high-throughput biological data analyses. He was accustomed to working with large data sets and biological modeling.

Trupti Joshi, an assistant computer science research professor, soon joined Xu’s research group to further support the center’s informatics goals. With a background in both biology and computer science, Joshi had a 180-degree view of plant biotechnology.

Joshi has had a passion for programming since 2000, when bioinformatics was a new field. “It was rapidly evolving field with cutting edge ideas. I found it so satisfying.”

In 2010, all four were members of an extended international team that successfully mapped soybean’s genome, a huge undertaking that generated a large amount of valuable genetic data. Making the data accessible to the larger science and agricultural communities was made possible by the National Science Foundation 2008 establishment of the iPlant Collaborative — cyberinfrastructure that anticipated the need for the supercomputing capabilities necessary for bioinformatic research — and the Obama Administration’s $200 million Big Data Initiative.

Xu and Joshi worked together initially to generate ideas for what they named the Soybean Knowledge Base (SoyKB). As lead designer and developer, it became Joshi’s dissertation project, and she has spearheaded the effort to build it into a comprehensive web resource featuring tools to functionlalize integrated access to the original genomic data and a growing body of published genetic data that MU researchers and others from all over the world are generating.

“We have more than 400 registered users, including domestic and international, as well as [users from] academia and industry, who access both public and private data in SoyKB,” said Joshi. “Users can access the public data without any registration and login. We have more than 1,000 users who regularly visit SoyKB every month.”

According to Joshi, one of the problems with contemporary soybean varieties is that through selective breeding, natural variability has been reduced, and one of the project’s goals is to bring in data from regions with natural variations. The ongoing, USB-supported project to identify variations in 500-plus soybean lines through the academic and industry partnership among MU, Dow Agrosciences, Monsanto and Bayer will generate extremely valuable datasets in this regard.

“Users can utilize their own data, which is one of the biggest benefits,” said Joshi of SoyKB’s architecture, which includes database, genome browser, web interface and data integration modules. “They can look at visualizations of how gene expression might be different in different situations and are able to take all of the datasets and go in and make decisions by looking at regions of interest to design a crop computationally. It cuts down on the [traditional crossbreeding] time to make the combinations.”

Nguyen identified genetic traits of key interest to soybean researchers as drought tolerance, stress tolerance, increased yield and disease resistance, adding that those with interest in soy as biofuel would like to increase soybean oil content. Soybeans are also being examined for their potential role in the prevention of cancer and other human diseases. Scientists working with food

Screenshot: Heatmap and hierarchical clustering analysis tools in SoyKB show expression of a handful of genes in soybean in various tissues.
grade soybeans would like to increase protein content and cooking oil quality, which is one of Nguyen’s interests.

“Researchers are able to use SoyKB to look at the gene structure and mutations to bring the right genetic combination of soybeans together,” Nguyen said. “Currently, soybean oil is 25 percent oleic acid, its most healthy component. We are working to make it 80 percent. It’s a game changer. Once in a while you get a breakthrough, and this is a breakthrough.

“Very soon, consumers in the U.S. can buy soybean oil that is healthier. It will revolutionize the use of oil in this country and the world,” he predicted.

Nguyen said the biotech research he is doing can be described as molecular breeding — using gene structure variation by changing a chemical base in a nucleotide sequence. Computational biology gives access to this marker-assisted selection, or breeding of natural variation traits, creating “better” genetics, not foreign genetics.

“Biotech [also] includes recombinant DNA technology [creating sequences that would not otherwise be found in biological organisms], which results in GMO crops, but not all biotech [results] are GMOs,” he explained in order to differentiate the work he is doing from genetic modification, which is viewed by some consumers as a controversial practice.

Joshi said their work has been supported by efforts from other computer scientists on the MU campus to expand the university’s cyberinfrastructure. National Science Foundation grants awarded to computer science Assistant Professor Prasad Calyam and Electrical and Computer Engineering Department Chair Chi-Ren Shyu were used to build a secure hybrid cloud networks, which support the high-volume data movement necessary for scientific research.

“It’s really exciting,” Joshi said. “It will give MU a high profile and gives us an opportunity to really make a difference. [At MU] there is the opportunity to work with so many groups. It’s interesting and rewarding.”
SoyKB also serves as an excellent tool and resource for training the next generation of plant breeders, computational biologists and computer scientists (see below).

Joshi recently has transitioned to the position of director of translational bioinformatics with the MU School of Medicine and will serve as assistant research professor in the Department of Molecular Microbiology and Immunology, but development of SoyKB will continue under her direction, through her many collaborations as a core faculty of MU Informatics Institute (MUII) and Interdisciplinary Plant Group (IPG).

“There is no way we could do this without the partnership between plant sciences and computer science,” Nguyen said of the collaboration. “SoyKB is a good bridging between CAFNR (College of Agriculture, Forestry and Natural Resources) and the College of Engineering.

“We have outstanding scientific recognition. It represents a lot of passion on the part of a lot of people.”

SoyKB project provides student researchers real world experience
by Jan Wiese-Fales

By design, computational projects used in undergraduate computer science courses generally are small-scale, idealized assignments, known as “toy problems,” that teach students how to respond to a hypothetical need. But scalability may remain elusive until students are able to apply the knowledge in a setting that supports large numbers of clients or patrons.

“This is really an important issue for students,” said Professor Dong Xu, department chair in MU’s Computer Science Department. “SoyKB provides a platform for students to work on real world applications with large-scale software engineering, enterprise-level software and the ability to work with and communicate with others in the profession.

“Companies are not in a position to train students, so it is our responsibility. [SoyKB] has produced some very successful students who have experience when they go into industry.”

Two students who got a taste of computational scalability and the other benefits of involvement with SoyKB are Michael Fitzpatrick and Danny Franklin.

Fitzpatrick, now a master’s student at Stanford University, currently is focused on genomics for a project/system known as DeepDive that uses machine-learning techniques for a variety of applications. He came to the MU campus in 2010 as a Discovery Fellow with the MU Honors College, guaranteeing him the opportunity to work with a faculty member on a research project.

“A few weeks into my freshman year, I started working in Dong Xu’s lab,” said Fitzpatrick, who started doing basic programming and worked with a web design tool when he was in the third grade. “A lot of the programming languages they were using were familiar, and I was able to make some immediate contributions.”

Franklin, who has worked as an operating systems software engineer at Microsoft Corp. for two years, began working with the SoyKB team in 2009 during his sophomore year.

“When I started with SoyKB, I didn’t know much about it at all. I’m surprised they even let me work there,” Franklin said. “[Assistant CS Research Professor] Trupti Joshi was very helpful and guided me through it. Then later, when I took [programming] classes, they were a breeze.”

Fitzpatrick said search engine analysis tools must be efficient to process all of the data and then the data must be displayed in a smart way. Visualization tools were one of the things he worked on for the project.

“SoyKB allows you to view a chromosome as well as gene expression because of the visualization tools,” Fitzpatrick said. “One of the reasons bioinformatics is so cool is that there are discoveries still being made and exciting problems to solve.”

Fitzpatrick has committed to work as a software engineer for Google when he graduates. Franklin is an operating systems software engineer at Microsoft. Photos courtesy Fitzpatrick and Franklin.